

Marketplace for AI Models

Abhishek Kumar

Department of Computer Science

University of Helsinki

abhishek.kumar@helsinki.fi

Benjamin Finley

Department of Computer Science

University of Helsinki

benjamin.finley@helsinki.fi

Tristan Braud

Department of Computer Science

HKUST

tbraud@cse.ust.hk

Sasu Tarkoma

Department of Computer Science

University of Helsinki

sasu.tarkoma@helsinki.fi

Pan Hui

Department of Computer Science

University of Helsinki & HKUST

pan.hui@helsinki.fi

ABSTRACT

Artificial intelligence shows promise for solving many practical societal problems in areas such as healthcare and transportation. However, the current mechanisms for AI model diffusion such as Github code repositories, academic project webpages, and commercial AI marketplaces have some limitations; for example, a lack of monetization methods, model traceability, and model auditability. In this work, we sketch guidelines for a new AI diffusion method based on a decentralized online marketplace. We consider the technical, economic, and regulatory aspects of such a marketplace including a discussion of solutions for problems in these areas. Finally, we include a comparative analysis of several current AI marketplaces that are already available or in development. We find that most of these marketplaces are centralized commercial marketplaces with relatively few models.

1 INTRODUCTION

Artificial intelligence (AI) is predicted to have a major societal impact over the coming decades. Specifically, when widely diffused, AI models have the potential to solve many ubiquitous problems in domains ranging from healthcare to transportation. For example, neural network based models have shown human and super-human level performance in many health related diagnostic tasks such as breast cancer detection [1].

However, to achieve widespread diffusion of AI models and thus capture these benefits, an efficient diffusion mechanism is required. Unfortunately, many popular diffusion mechanisms such as collections of Github repositories, academic research project pages, and existing commercial AI marketplaces have major limitations. Github repositories and academic research projects generally do not have straightforward monetization methods and installing, configuring, applying, and supporting models from these sources can be cumbersome (as such concerns are usually not paramount to academic researchers). While existing commercial AI marketplaces (refer to Section 7) are often centralized and controlled by a single company that may have different motives from the majority of the marketplace users thus allowing for conflict and a single point of failure. Additionally, the models available on commercial AI marketplaces may lag behind the state-of-the-art models available on Github or project pages. Finally, the datasets for training such AI models often cannot be provided for privacy or other reasons, thus the traceability and auditability of models from these sources

is lacking. Traits such as these will be important given new AI regulations currently in development by, for example, the European Union (refer to Section 6).

Given the limitations of existing mechanisms, in this paper we aim to sketch guidelines for a new AI diffusion mechanism based on a decentralized online marketplace and hereafter known as **AI marketplace**. We consider the technical, economic, and regulatory aspects of creating such a marketplace in order to reach the goal of broad yet ethical AI diffusion (as shown in Figure 1). The AI marketplace we propose would bring together various actors, including developers and companies of different sizes, users, data collectors, and even governmental entities, towards this common goal.

2 WHAT IS AI MARKETPLACE?

In our vision, the AI marketplace should facilitate two high level operations: 1) a developer should be able to sell their pre-trained AI models in the marketplace, and 2) a customer should be able to request a custom AI model which suits their specific needs, and the marketplace should be able to match such customers with developers who could build such a model. Also the marketplace should facilitate, when necessary, data sharing (selling and buying) and participation of marketplace actors in, for example, federated learning in support of the mentioned operations.

In a high-level sense, an AI marketplace is similar to other online marketplaces like eBay in terms of business operations and dynamics (e.g., two-sided network effects). However, at the same time, such a marketplace differs from those markets because of the nature of the products. Specifically, an AI marketplace is different from a conventional online marketplace in the following ways:

- Developing AI models often requires the sharing of data from the customer side, and such data may be proprietary. Therefore, an AI marketplace may have a mechanism which ensures that developers use that data only for training purposes.
- An AI marketplace needs a mechanism which can determine the quality of a final delivered AI model. Conventionally, accuracy has been the primary metric. However, alternative metrics that capture reliability, robustness, and fairness are also now considered important.
- Like in any conventional software system, AI systems also require maintenance over time. In the standard software

industry, the company which originally developed the software is usually responsible for providing support. However, in an AI marketplace, specific AI developers may not be available in the future. So, an AI marketplace needs standard guidelines which AI developers should follow while developing models for marketplace customers. Thus maintenance by other AI developers is much easier.

An AI marketplace could also be considered similar to a mobile app store like Google Playstore and Apple App Store. These stores host many AI-enabled applications, however, they still differ from an AI marketplace in some significant ways:

- Unlike in an app store, an AI marketplace could allow customers to request new products on the fly. An AI marketplace can then quickly match AI customers with AI developers with relevant expertise.
- Some AI models are proprietary. So, unlike in an app store, they can not be shared online with a wider audience, as this would further facilitate adversarial attacks and risk leaking intellectual property.

Finally, an AI marketplace is also similar to an online data marketplace. For example, in both marketplaces the products sold could leak private user data. Specifically, attacks against AI models like model inversion [12][46] and membership inference [35][40] can extract information about users whose data was used in training. However, data marketplaces can still be considered closer to conventional online marketplaces, rather than AI marketplaces, since the transferring of data can be considered analogous to the transferring of physical products (assuming that privacy regulations have been followed) as after the transfer (or successful trade) little maintenance is needed, unlike for AI models.

3 DRIVING FORCES BEHIND AI MARKETPLACE

An AI Marketplace aims to respond to several issues within the current AI community. Specifically, in order for AI to diffuse and achieve widespread adoption, it is necessary to address the following concerns:

Lack of interoperability standards: Currently, there are multiple frameworks for developing diverse AI models. Different developers use different frameworks (TensorFlow, PyTorch, Caffe2), different languages (Python, Java, C/C++), and target different environments (powerful Linux server, smartphone, minimalist IoT device) depending on the intended usage. Each of these elements comes with challenges. A small AI company may want to pipeline a server-based PyTorch AI model with an externally developed smartphone-based TensorFlow AI model. However, the current lack of interoperability standards dramatically limits such opportunities, and adapting existing models is a tedious task (potentially including redeveloping an application-specific AI model) [13][25].

Lack of infrastructure for AI data cooperation: Many state-of-the-art AI models, especially those based on deep learning, require very large datasets. Unfortunately, the creation, management, and sharing of very large datasets is often difficult for many AI developers (due to resource or capability limitations). As a result, AI development with large data is dominated by researchers in large organizations that have significant capabilities and resources.

Rise of data protection regulations: Collecting user's data is increasingly difficult due to privacy regulations around the world, such as the European General Data Protection Regulation (GDPR) or the California Consumer Privacy Act. These regulations put the burden of protecting the user's privacy on the shoulders of data collectors. Users may give consent to collect their personal data at a given point in time, but they can also withdraw their consent later. The data collector is then often required to erase the collected data. Furthermore, fulfilling these requests requires both technical expertise and regulatory expertise. Most AI developers do not possess both of these skills [33].

Large cost of AI development and operation: According to a Teradata survey, the lack of IT infrastructure (40%) and the lack of talent (34%) are the two most significant barriers to AI realization [43]. The current lack of qualified AI professionals makes it expensive to hire an AI team. A fully-fledged AI team not only consists of AI developers, but also domain experts, data engineers, product designers, AI sociologists, and IT lawyers. Most small businesses can not afford to hire such an expensive team [51]. Additionally, as mentioned, the infrastructure to collect and store the massive amounts of data required for model development, training, and operation is also costly.

An AI marketplace is a potential solution for overcoming many of these barriers. An AI marketplace can make AI models and datasets accessible to end-users and developers, and give developers a way to monetize their models.

4 TECHNICAL ASPECTS OF AI MARKETPLACE

In a basic AI marketplace setting, an AI customer may arrive with a training dataset, and another smaller dataset referred to as a "validation" dataset, and want to build a prediction model that performs well on this validation dataset. The AI marketplace can match the customer with AI developers with the skills needed to build a model. If the model developed by the AI developer based on the training dataset (provided by the customer) achieves sufficiently high accuracy (or other metric(s) with a threshold set by the customer) on the validation dataset, then customer and developer can move forward with the transaction.

However, the potential AI model (e.g., a deep neural network like ResNet or Inception) may require a large amount of training data which the customer may not have. Furthermore, the original training dataset provided by the customer may itself be from multiple private sources (e.g., mobile crowdsensing, shopping patterns) and may follow some multimodal distribution. The AI marketplace should also be able to help the customer by allowing the aggregation of multiple alternative datasets from the marketplace while also ensuring the aggregate dataset follows a similar distribution (e.g., using a transfer learning approach [50]) as the validation dataset. This is important since AI learning algorithms suffer from major model quality loss (or even divergence) when trained on non-IID data [19]. In the process, the marketplace should also enable the monetization of data in a trusted, fair manner while preserving data ownership and privacy as much as possible [29].

Let us consider the following scenario in the healthcare domain. The consumer is a newly established cancer treatment hospital, and

the data sources are cancer institutions from different geographical locations across the globe. The goal of the new hospital is to construct an ML model that can predict the early onset of a given form of cancer. The model must perform well given the demography of its patients, and therefore it is crucial to collect data similar to the small validation set that is representative of the demography. However, the individual data sources have widely different demographics data due to their locations. The goal of an AI Marketplace in such a setting is to enable the collection of a dataset sampled from these sources that matches the demography of the new hospital and in the process attributes fair value to the different data sources. In other settings, the data owner or consumer may not have sufficient AI expertise or skills. The marketplace should be able to connect the consumer to AI experts, and at the same time, should provide a platform for AI experts to assess the performance of their models on the consumer's validation dataset without direct access to the dataset (by the AI experts).

A model/data exchange mechanism in an AI marketplace should have the following properties:

- All individual data sources and the consumer should have privacy protection in the form of differential privacy guarantees to communicate with a specific actor (data owner, AI expert, customer). Such differential privacy guarantees can be achieved by ensuring that all data transferred (between the aggregator and various data owners) is entirely anonymized, de-identified, and ideally encrypted.
- Transfer learning should be facilitated in the sense that the aggregating entity acquires a summary training dataset that is statistically related to the consumer validation dataset with respect to the requested metrics.
- The aggregator can only learn the pairwise Euclidean distances between the points in the training dataset.
- Consumers having sufficient AI skills, but no or little training data should be able to leverage federated learning on the marketplace to facilitate learning without moving raw training data from its original owners. However, to facilitate quick convergence, the marketplace should locate data sources which are nearly identical and independently distributed.
- All entities, i.e., AI model contributor or data contributor, should be awarded fairly for their contribution. Unlike normal goods in online marketplaces like eBay, deciding the fair award (or data/model price) is non-trivial in an AI marketplace.
- No collaborating entities, i.e., data provider or model training provider in case of federated learning, should be able to cheat in the model building process. In other words, the marketplace should provide robustness against data providers or training update providers who commit malicious or low quality contributions. Thus the marketplace should incorporate a verification mechanism which will assess the quality of the contribution, and hence will determine rewards accordingly.

In addition to the properties mentioned above, AI marketplace should also incorporate other relevant properties from conventional marketplaces, e.g., ensuring liquidity in the market [48], providing a

framework for conflict resolution between consumer and seller [44]. In the remainder of this section, we discuss potential solutions to address these properties.

4.1 Maintaining data privacy in AI marketplace

In an AI marketplace context, two major strategies can ensure data privacy for the involved actors.

Federated learning/Peer-to-Peer (P2P) learning as learning paradigms: Under the federated learning framework [21][17] or the P2P learning framework [3][8], raw data never leaves the owners' devices, and thus data privacy and ownership of the original raw data is always ensured with data owners. However, such paradigms add computational overload in the form of local training on the data owner's end.

Using contextual integrity as design principle for data sharing: The idea of privacy trading is also gaining popularity [4][20]. Many users who are less privacy-sensitive may be willing to sell their privacy. However, even such users should know the future use of their raw data before selling it. The principle of contextual integrity allows enforcing data privacy by providing a framework for evaluating the flow of personal information between different recipients and explaining why certain patterns of information flow are acceptable in one context but problematic in another [36][54]. The contextual integrity framework allows users to maintain control of their data even after trading it.

4.2 Price determination in AI marketplace

Ensuring that the marketplace is transparent in assigning value based on the quality of the data or model relative to the target is extremely important. Unlike in conventional online marketplaces, deciding the right price of data or an AI model is non-trivial. In conventional online marketplaces, the price in the offline market can serve as a reference. Also, in a corporate AI setting, experts can estimate the resources required for a project and hence negotiate the price of AI models with potential buyers. However, in an online AI marketplace, an individual AI developer may not possess the skills to determine the right price for their model.

To ensure fairness when deriving the price, the marketplace should provide a bidding mechanism. This bidding mechanism should ideally 1) be a dominant strategy and incentives-compatible. The dominant strategy of each entity is to bid the amount equal to their private valuation. Bidding this true valuation always leads to a non-zero utility for any entities, 2) maximize the social surplus when all entities report their valuations truthfully, and 3) be implementable in polynomial (preferably linear) time in order to enable scalability. Any mechanism satisfying these three conditions can be said to have employed Vickrey auction [47].

Another approach could be to organize contests among AI developers for some pre-defined reward. However, contests and their corresponding reward incentives should be designed based on accurate models of AI developers' strategic behaviour to elicit the desired outcomes [14]. Depending on the strategic behaviour of the AI developers, different kinds of contests can be organized, e.g., contests which reward a fixed number of AI developers, contests which take the form of a tournament, or contests which award everything to the winner [9][15].

4.3 Robustness against malicious entities in AI marketplace

A Federated learning or P2P learning framework ensures data privacy and ownership of the data to the original owners. However, such a framework also introduces a number of vulnerabilities. Some entities may want to free-ride by trying to capture benefits (or payments) without making honest contributions, e.g., sending random training updates to the server instead of updates calculated on real data after local training. In other scenarios, other competitors/adversaries may try to introduce model poisoning in order to reduce the reputation of the given marketplace. An AI marketplace should have a verification mechanism to assess whether the given data or training updates are coming from free-riders [27], malicious entities [5], or honest users.

4.4 Auditability in AI marketplace

The entire process must be transparent and immutable in order to ensure trust and fairness. Based on how this process is enforced in the marketplace, there exists two types of marketplaces: 1) centralized marketplaces, where a trusted entity ensures smooth operations and maintains an immutable log of all operations on the platform, and 2) decentralized marketplaces, where no single entity is solely trusted by one entity. Instead, all operations are stored on an immutable public distributed ledger (or a public Blockchain). Both marketplace types have their pros and cons. In a centralized marketplace, since the central managing entity earns revenue by charging a small transaction fee for each successful transaction, the entity is motivated to maintain smooth operations on the marketplace by verifying the identity of all parties, improving matching mechanisms, supporting buyer-seller conflict resolution, and ensuring liquidity in the market. In a decentralized AI marketplace, all these goals can be achieved, at the expense of higher energy and time consumption since any changes in the policy must have consensus from peers on the marketplace. Current consensus algorithms are energy-intensive [2], but these are likely to improve in the future. A decentralized AI marketplace is a true enabler of the vision of democratization of the current tightly concentrated AI ecosystem among a limited number of large players.

5 ECONOMIC ASPECTS OF AI MARKETPLACE

Online marketplaces represent an interesting business model in that they facilitate transactions between suppliers and customers often without taking possession of or full responsibility for products or services; thus they have very low cost structures and very high gross margins (e.g., 70% for eBay, 60% for Etsy). Additionally, network effects make them highly defensible. For example, Alibaba, Craigslist, eBay, and Rakuten are more than 15 years old, but still dominate their sectors. In the past ten years, the number of online marketplaces worth more than \$1 billion has gone from two (Craigslist and eBay) to more than a dozen in the United States (Airbnb, Etsy, Groupon, GrubHub Seamless, Lending Club, Lyft, Prosper, Thumbtack, Uber, and Upwork). That number is expected to double by the end of 2020, according to Greylock Partners, a Silicon Valley venture capital firm [16].

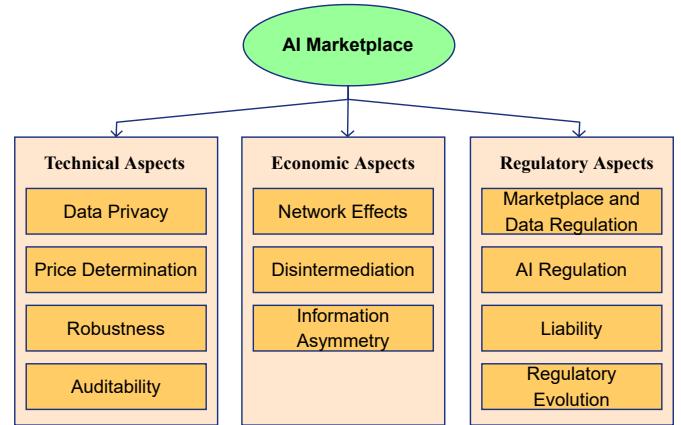


Figure 1: Aspects of an AI Marketplace

In order to build a successful AI marketplace, a critical number of buyers and suppliers of AI models are needed, just like in any other online marketplace. Potential suppliers for such a marketplace could be AI developers of algorithms and software on platforms like Github, whereas potential buyers could be companies that can not afford their own team of AI experts.

5.1 Network effects

A network effect is a phenomenon whereby the value of a platform or service (to an individual participant) is proportional to the number of participants. This phenomenon is often the single-most important factor behind the success or failure of any online platform. Platforms like eBay and Facebook continue to dominate their respective markets partly because they exploit these network effects very well, whereas platforms like Google Plus did not take off partly due to being on the wrong end of such network effects [53][23]. So, network effects would be a crucial element for the success or failure of an AI marketplace as well.

The current AI ecosystem favors major players like Google and Facebook as they are already exploiting so called data network effects very well. Specifically, AI-based products or services from these companies become smarter as they train on more data from more users [34][26]. In turn, smarter products and services attract more new users, thus creating a feedback loop.

With the adoption of two-tier model training architecture like federated learning [17] (e.g., the first tier being a general global model and the second tier being a personalized local model), challenging the dominance of these companies will be even more difficult. Specifically, this architecture allows these large companies to provide personalized model training to AI customers. The companies provide a pre-trained general model (trained on datasets either owned by the company or procured by the company), and the AI customer personalizes by training on their local dataset. This training paradigm works well even with smaller amounts of a user's training data.

So, if an AI marketplace wishes to challenge the dominance of these companies, it needs to support interoperability between

datasets and models so that smaller developers can efficiently aggregate enough smaller training datasets, federated training users, or even models (into an ensemble) to compete (or at least reach a minimum threshold) in terms of performance.

5.2 Disintermediation

Conventional online marketplaces fear that once they facilitate a successful transaction, the buyer and the seller will agree to conduct their subsequent interactions outside the marketplace. However, such a risk could be minimized for an AI marketplace. Specifically, many AI models tend to be black box in nature and an AI marketplace could offer buyers independent assessments of the AI model which they are about to procure (thus providing an added-value service). Such assessments might even be mandated by future government regulations as discussed in section 6.

Other strategies could also incentivize AI customers to deal with AI developers through the marketplace. For example, an AI marketplace could also ensure the prices for products without added value (like the mentioned assessments) are the same as when consumers buy directly from the seller or via other channels, otherwise this may also lead to a disintermediation problem.

5.3 Information asymmetry

In terms of matching AI customers with AI developers, an AI marketplace is an online freelance marketplace that matches buyers of electronically deliverable services with freelancers. Just like in any freelancing marketplace, AI customers may also face the issue of “information asymmetry”, i.e., they may face uncertainty over the quality of individual AI developers. A solution to this dilemma is a trust or reputation mechanism to help facilitate transactions between strangers. Unlike in a conventional online marketplace where transactions mostly involve products/services and monetary payments, transactions in an AI marketplace may also involve training datasets, which may themselves have economic value; therefore mechanisms of an AI marketplace should ensure an even greater degree of trust. Luca et al. [28] found that in an online freelance marketplace, customers are forward-looking and that they place significant weight on a seller’s reputation. Though, not controlling for buyers’ inter-temporal trade-offs and dynamic selection can considerably bias such reputations. Thus, an AI marketplace should not rely entirely on a reputation mechanism built on reviews from buyers and sellers to tackle the issue of information asymmetry [52]. The marketplace needs to play an active role to ensure fairness in these reputations mechanisms. Given that many AI models are black box in nature, the marketplace may need to support independent assessments of an AI model (before final procurement) as suggested in Section 4, reputation mechanisms could be primarily centered around such assessments in order to ensure fairness.

6 REGULATORY ASPECTS OF AI MARKETPLACE

The regulation of AI marketplaces, as well as the regulation of AI in general, is still a significant unknown with major countries only beginning to grapple with the difficult task. The regulation of any AI marketplace as such would combine the regulatory frameworks from several different domains: regulation of the often sensitive

training/testing data, regulation of the application of the AI model (often in sensitive domains), and regulation of online marketplaces.

6.1 Market and data regulation

The regulation of online marketplaces and sensitive data have historical precedents (e.g., Sherman Act and Health Insurance Portability and Accountability Act (HIPAA) in the USA) due to the analogues of offline marketplaces and physical data. These regulations are thus being overhauled (e.g., GDPR in Europe) to better deal with the issues brought by the internet and AI eras. Therefore, an AI marketplace must comply with these new regulations. As discussed, the fundamental architecture of an AI marketplace can help ensure compliance through data privacy mechanisms such as differential privacy and federated learning and prevention of marketplace monopoly through open interoperability standards (no lock-in and low switching costs). Unfortunately, even current regulations like GDPR are vague in many cases, thus creating uncertainty and cautiousness in companies that may deter participation in an AI market. These uncertainties are compounded by the differing approaches to such regulations in varying markets [39]. We discuss this topic further in Section 6.4.

6.2 AI regulation

In terms of AI itself, the regulation is currently very sparse but developing. For example, the EU recently released a draft paper outlining its vision for AI regulation in high-risk areas (e.g., transportation or health care) [11][49], as well as ethical guidelines for building trustworthy AI systems [18][22]. The draft includes an overarching framework that would cover the training data (including data traceability and coverage to ensure fairness), model transparency (to understand why certain decisions are taken), and liability (in case of harm).

Similar to the data-centric regulations, an AI marketplace can ensure that at least some subset of the marketplace (e.g., a high-risk area section) enforces or checks that the AI follows these regulations. For example, data traceability can be ensured (at least up to the point of individual marketplace actors) through the use of the public ledger to track the actors providing the data [42]. Similarly, adequate training data coverage (to prevent discrimination or bias by AI [37]) can be ensured through the use of diverse third party (e.g., even government or conformity bodies) validation datasets enforced in a smart contract. In fact, the EU already discusses in the draft paper the potential for “support structures” and “online tools [that] could facilitate compliance” to help especially small and medium size businesses [11]. The marketplace could also periodically re-verify compliance as models evolve as such regulations apply both ex-ante and ex-post. As a final example, data privacy in the visual domain could be ensured by incorporating a mandatory privacy-respecting mechanism in vision-based applications [41].

Interestingly, in the US, for example, the healthcare domain does have some regulations for AI/ML models partly derived from existing regulations on healthcare software. In fact, many of the models available on existing commercial AI marketplaces are healthcare based (refer to Section 7). As an example of a current regulatory problem, the US regulator is discussing how to regulate AI/ML models that frequently or continuously learn without requiring a

regulatory review after every model update (which could be the case under current regulations) [45]. Again, such future regulations could potentially leverage automated testing on independent government or conformity body validation datasets through an AI marketplace.

Even unconventional organizations are delving into the area, with the Vatican organizing a workshop denoted as “The ‘Good’ Algorithm? Artificial Intelligence: Ethics, Law, Health” [31]. Additionally, AI regulation and governance has been the subject of recent interdisciplinary academic research by computer scientists, lawyers, and others [7][30].

6.3 Liability regulation

Liability in the case of such an AI marketplace is also a difficult problem stemming from the difficulty of liability in both AI and online marketplace platforms contexts. Hereafter, given the economic context, we focus on civil liability as opposed to criminal liability.

Firstly, civil liability, in general, must balance the need to incentivize product safety and compensate victims of harm with the need to encourage business innovation. This balance is especially difficult given the rapid innovation in AI and the potential economic and societal benefits of AI. Additionally, in many legal systems, such as the EU, for compensation the victim must prove damage, a product defect, and a causal link between the two [10]. With complex AI or software-based systems, identifying the liable person can be burdensome or in cases with human-AI collaborative systems the liable person may be unclear. A possible solution is to alter the burden of proof requirement, for example, by inverting the burden to rest with the producing company. This inversion would require companies to have very clear and coherent tracking and documenting of AI models which the AI marketplace inherently enables.

In terms of online marketplaces, the liability of companies such as Amazon for products from third-party sellers on their platform (about 58% of Amazon sales) is a matter of ongoing legal debate [24][6]. For example, in the US, the issue of liability currently revolves around whether Amazon is considered in a legal sense “a seller” or simply “a platform for sellers and buyers”. This, in turn, is primarily related to how much power they have over the third party sellers (along with several other considerations). Court cases (e.g., Oberdorf v. Amazon) are currently in progress, and a case may eventually reach the US supreme court. Currently, the status quo in the US is that amazon is not liable. The situation in Europe is similar, with on-going work on developing new regulations and eventually adapting the EU Product Liability Directive [6]. Given this background, under current trends, if the AI marketplace does not exert strict control and gain excessive power over sellers, then liability could be minimized by maintaining the status of a platform.

6.4 Regulatory evolution

Overall, any AI marketplace would need to evolve along with novel regulation (e.g., safety or export regulations) or risk becoming unusable by legitimate users. Specifically, several distinct marketplaces or strict marketplace access controls may be necessary given new

export regulations. For example, new US regulations require companies to have a special license to export certain geospatial AI software [38]. The justification for the new regulations is based on national security (with especially China in mind [32]). Additionally, a marketplace may need to follow the strictest common safety regulations given the potential for safety regulatory divergence between the US, EU, China, and others. For example, with data privacy, many global internet companies are now GDPR compliant even if they are primarily domiciled elsewhere because they have European interests or customers.

Unfortunately, even in the long term, regulatory convergence may be difficult because AI is also viewed as a strategic security and economic asset to many countries, and thus some do not want to impede any technological progress with regulation [39].

7 WHERE WE ARE NOW?

In Table 1, we provide the list of companies which are either providing an AI marketplace or in the process of building such a marketplace. To find such companies, we searched on Google using the keywords: “online marketplace”, “data marketplace”, and “AI marketplace” and extracted the first ten pages. We then manually visited all links from these pages and checked which describe an entity providing for the trading of AI models, a service to enable trading of AI models, or are in the process of building either of these. After six pages, the search results no longer provided any meaningful links. Eventually, we identified the 24 companies or frameworks listed in Table 1.

Most of these companies are based in either the USA or Europe. Among those which are currently available, none list more than 24 different models (Nauance Communications lists 24, Gravity AI lists 12, and IBM Imaging lists five models), thus suggesting that none have seen major or widespread adoption. The marketplaces which are somewhat mature primarily focus on the healthcare domain. However, these marketplaces are not operating in multiple countries potentially due to the need for regulatory approval of such models in each country or economic area (for example, by the US Food and Drug Administration).

The overarching goal of most of these AI marketplaces aligns with our vision of a general marketplace where buyers and sellers engage in transactions for AI models. In terms of the technical aspects for a successful AI marketplace, from Section 4, most of the marketplaces have not yet incorporated these though they do often acknowledge the need for such aspects. For example, only two of them mention that they support scalable privacy-preserving model training paradigms like federated learning. As for architecture, most of the marketplaces are proprietary and are based on a centralized architecture. Though the few decentralized marketplaces are primarily based on distributed ledgers, similar to our vision. With regard to the pricing mechanism, most of them are using fixed pricing per model.

8 CONCLUSION

In this work, we outlined the principles for a marketplace for AI models based on a decentralized online structure. Such a marketplace could help democratize and diffuse AI technology to the

Table 1: Existing AI marketplace frameworks and implementations

AI Marketplace	Domain Focus	Org Type	Architecture	Target Market	Domicile	Status
Nuance Communications	Diagnostic Imaging	For Profit	Centralized	USA & Canada	USA	Online
Agorai	Various ^a	For Profit	Centralized	Global	Singapore	Online
IBM Imaging	Healthcare	For Profit	Centralized	USA	USA	Online
Envoy AI	Medical Imaging	For Profit	Centralized	USA	USA	Online
GraphGrail AI	Various ^b	For Profit	Centralized	Russia	Russian	Online
Algorithmia	General	For Profit	Centralized	USA	USA	Online
Neuromation	Various ^d	For Profit	Centralized	Global	USA	Online
Ocean Protocol	General	For Profit	Decentralized	Global	Singapore	Online
OVHcloud AI	General	For Profit	Centralized	Global	France	Beta
Orange AI	General	For Profit	Centralized	Global	France	Beta
Kynisys	Various ^c	For Profit	Centralized	Global	UK	Beta
Gravity AI	General	For Profit	Centralized	Global	USA	Beta
SingularityNET	General	Non-profit	Decentralized	Global	Netherlands	Beta
Modzy	General	For Profit	Centralized	USA	USA	Alpha
Alphacat	Fintech	For Profit	Centralized	Global	-	Alpha
Bonsneys	General	For Profit	Centralized	Europe	Switzerland	In-development
Akira AI	General	For Profit	Centralized	Global	India	In-development
Genesis AI	General	For Profit	Centralized	Global	USA	In-development
AI Global	General	Non-profit	Centralized	Global	USA	In-development
Synapse AI	General	For Profit	Decentralized	Global	USA	In-development
TensorTask	General	Non-profit	Decentralized	Global	USA	In-development
Nomidman	General	For-profit	Decentralized	Global	Estonia	In-development
OSA Decentralized	Various ^e	For-profit	Decentralized	Global	BVI ^g	In-development
DaiMoN	General	Non-profit	Decentralized	-	-	PoC ^f

^a Finance, Healthcare, Retail, & Advertising

^b Finance, Travel, Retail, Advertising, & Consumer Goods

^c Security & IoT, Oil & Gas, Robotics, & Automobile

^d Surveillance, Retail, Medical Imaging, Industrial Robotics, & Manufacturing

^e Retail, Manufacturing, & Consumer Goods

^f Proof of Concept (Academic Work)

^g British Virgin Islands

greater society, including to smaller actors (like small and medium-size companies). We discussed the technical, economic, and regulatory aspects to consider while designing such a marketplace. We also described (often novel) technologies and solutions that can help address problems in these aspects. For example, utilizing federated learning for privacy-preserving machine learning across marketplace actors. Finally, we studied the current state of various AI marketplaces and provided a comparative analysis of these marketplaces based on properties such as architecture, domain, and status. We found that most of these currently available marketplaces are centralized and company-driven with relatively few models per marketplace. Thus suggesting that AI marketplaces are still in their infancy.

ACKNOWLEDGMENTS

This work has been supported by the 5GEAR project funded by the Academy of Finland (Decision No. 319669) and the FIT project funded by the Academy of Finland (Decision No. 325570).

REFERENCES

- [1] Mohamed Alloghani, Dhiya Al-Jumeily, Ahmed J. Aljaaf, Mohammed Khalaf, Jamila Mustafina, and Sin Y. Tan. 2020. The Application of Artificial Intelligence Technology in Healthcare: A Systematic Review. In *Applied Computing to Support Industry: Innovation and Technology*, Mohammed I. Khalaf, Dhiya Al-Jumeily, and Alexei Lisitsa (Eds.). Springer International Publishing, 248–261.
- [2] Michael Barnard. 2018. *Blockchain - An Innovation Enabler for Clean Technology*. Technical Report. <https://future-trends.cleantechnica.com/reports/blockchain-innovation-enabler-clean-technology-2018/>.
- [3] Aurélien Bellet, Rachid Guerraoui, Mahsa Taziki, and Marc Tommasi. 2018. Personalized and Private Peer-to-Peer Machine Learning. In *International Conference on Artificial Intelligence and Statistics*, 473–481.
- [4] Alastair R Beresford, Andrew Rice, Nicholas Skehin, and Ripudaman Sohan. 2011. Mockdroid: trading privacy for application functionality on smartphones. In *Proceedings of the 12th workshop on mobile computing systems and applications*, 49–54.
- [5] Arjun Nitin Bhagoji, Supriyo Chakraborty, Prateek Mittal, and Seraphin Calo. 2019. Analyzing Federated Learning through an Adversarial Lens. In *International Conference on Machine Learning*, 634–643.
- [6] Christoph Busch. 2019. When Product Liability Meets the Platform Economy: A European Perspective on Oberdorf v. Amazon. *Journal of European Consumer and Market Law* 8 (2019), 173–174.
- [7] Corinne Cath. 2018. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2133 (2018), 20180080. <https://doi.org/10.1098/rsta.2018.0080>

- [8] Hsin-Pai Cheng, Patrick Yu, Haojing Hu, Feng Yan, Shiyu Li, Hai Li, and Yiran Chen. 2018. LEASGD: an Efficient and Privacy-Preserving Decentralized Algorithm for Distributed Learning. *CoRR* abs/1811.11124 (2018). arXiv:1811.11124 <http://arxiv.org/abs/1811.11124>
- [9] David Easley and Arpita Ghosh. 2015. Behavioral mechanism design: optimal crowdsourcing contracts and prospect theory. *ACM SIGecom Exchanges* 14, 1 (2015), 89–94.
- [10] European Commission. 2020. *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*. Technical Report. <https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics.en>.
- [11] European Commission. 2020. *White Paper on Artificial Intelligence A European approach to excellence and trust*. Technical Report. <https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust-en>.
- [12] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. 2015. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 1322–1333.
- [13] Ishaya GAMBO, Oluwatolani OLUWAGBEMI, and Philip ACHIMUGU. 2011. Lack of Interoperable Health Information Systems in Developing Countries: An Impact Analysis. *Journal of Health Informatics in Developing Countries* 5, 1 (2011).
- [14] Arpita Ghosh and Robert Kleinberg. 2016. Optimal contest design for simple agents. *ACM Transactions on Economics and Computation (TEAC)* 4, 4 (2016), 1–41.
- [15] Oliver Görtler and Matthias Kräkel. 2010. Optimal tournament contracts for heterogeneous workers. *Journal of Economic Behavior & Organization* 75, 2 (2010), 180–191.
- [16] Andrei Hagiu and Simon Rothman. 2016. Network effects aren't enough. *Harvard business review* 94, 4 (2016), 64–71.
- [17] Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. 2018. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604* (2018).
- [18] AI Hleg. 2019. Ethics guidelines for trustworthy AI. *B-1049 Brussels* (2019).
- [19] Kevin Hsieh, Amar Phanishayee, Onur Mutlu, and Phillip B. Gibbons. 2019. The Non-IID Data Quagmire of Decentralized Machine Learning. *CoRR* abs/1910.00189 (2019). arXiv:1910.00189 <http://arxiv.org/abs/1910.00189>
- [20] Wenqiang Jin, Mingyan Xiao, Ming Li, and Linke Guo. 2019. If you do not care about it, sell it: Trading location privacy in mobile crowd sensing. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 1045–1053.
- [21] Jakub Konečný, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. 2016. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492* (2016).
- [22] Abhishek Kumar, Tristan Braud, Sasu Tarkoma, and Pan Hui. 2020. Trustworthy AI in the Age of Pervasive Computing and Big Data. In *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, Austin, United States.
- [23] Marcel Landeweerd, Ton Spil, and Richard Klein. 2013. The success of Google search, the failure of Google health and the future of Google plus. In *International Working Conference on Transfer and Diffusion of IT*. Springer, 221–239.
- [24] Jacob Lehman and Mike Dolan. 2020. Strict Products Liability: One More Thing the Internet Is Disrupting. <https://www.law.com/thelegalintelligencer/2020/01/22/strict-products-liability-one-more-thing-the-internet-is-disrupting/>.
- [25] Moritz Lehne, Julian Sass, Andrea Essewanger, Josef Schepers, and Sylvia Thun. 2019. Why digital medicine depends on interoperability. *NPJ Digital Medicine* 2, 1 (2019), 1–5.
- [26] Shengli Li, Yipeng Liu, and Subhajyoti Bandyopadhyay. 2010. Network effects in online two-sided market platforms: A research note. *Decision Support Systems* 49, 2 (2010), 245–249.
- [27] Jierui Lin, Min Du, and Jian Liu. 2019. Free-riders in Federated Learning: Attacks and Defenses. *arXiv preprint arXiv:1911.12560* (2019).
- [28] Michael Luca. 2017. Designing online marketplaces: Trust and reputation mechanisms. *Innovation Policy and the Economy* 17, 1 (2017), 77–93.
- [29] John Lucker. 2015. The Dangers of Monetizing Data. <https://deloitte.wsj.com/cio/2015/06/30/the-dangers-of-monetizing-data/>.
- [30] Orla Lynskey. 2019. Criminal justice profiling and EU data protection law: precarious protection from predictive policing. *International Journal of Law in Context* 15, 2 (2019), 162ff?!176. <https://doi.org/10.1017/S1744552319000090>
- [31] Benedict Mayaki. 2020. Vatican Workshop on Ethics and AI. <https://www.vaticannews.va/en/vatican-city/news/2020-02/vatican-artificial-intelligence-ethics-workshop.html>.
- [32] Evan S Medeiros. 2019. The Changing Fundamentals of US-China Relations. *The Washington Quarterly* 42, 3 (2019), 93–119.
- [33] Vida Ahmadi Mehri, Dragos Ilie, and Kurt Tutschku. 2018. Privacy and DRM Requirements for Collaborative Development of AI Applications. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*. 1–8.
- [34] Hitoshi Mitomo et al. 2017. Data Network Effects: Implications for Data Business. In *28th European Regional ITS Conference, Passau 2017*. International Telecommunications Society (ITS).
- [35] Milad Nasr, Reza Shokri, and Amir Houmansadr. 2019. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 739–753.
- [36] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.
- [37] Irini Ntoutsi, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Ester Vidal, Salvatore Ruggieri, Franco Turini, Symeon Papadopoulos, Emmanouil Krasanakis, et al. 2020. Bias in Data-driven AI Systems—An Introductory Survey. *arXiv preprint arXiv:2001.09762* (2020).
- [38] Reuters. 2020. U.S. government limits exports of artificial intelligence software. <https://www.reuters.com/article/us-usa-artificial-intelligence-us-government-limits-exports-of-artificial-intelligence-software-idUSKBN1ZZ1PT>.
- [39] David Shepardson. 2020. White House proposes regulatory principles to govern AI use. <https://www.reuters.com/article/us-tech-ces-ai-white-house-white-house-proposes-regulatory-principles-to-govern-ai-use-idUSKBN1Z60GL>.
- [40] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 3–18.
- [41] Jiayu Shu, Rui Zheng, and Pan Hui. 2018. Cardea: Context-Aware Visual Privacy Protection for Photo Taking and Sharing. In *Proceedings of the 9th ACM Multimedia Systems Conference (MMSys'18)*. Association for Computing Machinery, New York, NY, USA, 304ff?!315. <https://doi.org/10.1145/3204949.3204973>
- [42] Jatinder Singh, Christopher Millard, Chris Reed, Jennifer Cobbe, and Jon Crowcroft. 2018. Accountability in the IoT: Systems, law, and ways forward. *Computer* 51, 7 (2018), 54–65.
- [43] Teradata. 2017. *State of Artificial Intelligence for Enterprises*. Technical Report. www.multivu.com/players/English/8075951-teradata-state-of-artificial-intelligence-ai-for-enterprises.
- [44] Meropi Tzanetakis, Gerrit Kamphausen, Bernd Wersle, and Roger von Laufenberg. 2016. The transparency paradox: Building trust, resolving disputes and optimising logistics on conventional and online drugs markets. *International Journal of Drug Policy* 35 (2016), 58–68.
- [45] U.S. Food and Drug Administration. 2020. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback*. Technical Report. U.S. Food and Drug Administration.
- [46] Michael Veale, Reuben Binns, and Lilian Edwards. 2018. Algorithms that remember: model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2133 (2018), 20180083.
- [47] William Vickrey. 1961. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance* 16, 1 (1961), 8–37.
- [48] Nir Vulkan. 2003. *The economics of e-commerce: a strategic guide to understanding and designing the online marketplace*. Princeton University Press.
- [49] Nicholas Wallace. 2020. Europe plans to strictly regulate high-risk AI technology. <https://www.sciencemag.org/news/2020/02/europe-plans-strictly-regulate-high-risk-ai-technology>.
- [50] Jindong Wang, Yiqiang Chen, Wenjie Feng, Han Yu, Meiyu Huang, and Qiang Yang. 2020. Transfer Learning with Dynamic Distribution Adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 11, 1 (2020), 1–25.
- [51] WebFX. 2020. AI Pricing: How Much Does Artificial Intelligence Cost? <https://www.webfx.com/internet-marketing/ai-pricing.html>.
- [52] Hema Yoganarasimhan. 2013. The value of reputation in an online freelance marketplace. *Marketing Science* 32, 6 (2013), 860–891.
- [53] Feng Zhu and Marco Iansiti. 2019. Why some platforms thrive and others don't. *Harvard Business Review* (2019).
- [54] Michael Zimmer. 2018. Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity. *Social Media + Society* 4, 2 (2018).